



文章栏目：“无废城市”建设专题

DOI 10.12030/j.cjee.202112202 中图分类号 X84 文献标识码 A

薛军, 奚强, 徐淑民, 等. 基于大数据方法的垃圾焚烧发电厂环境执法监管数据预测模型探究[J]. 环境工程学报, 2022, 16(3): 752-758.
[XUE Jun, XI Qiang, XU Shumin, et al. Prediction model analysis of environmental law enforcement supervision of waste incineration power plant based on big data method[J]. Chinese Journal of Environmental Engineering, 2022, 16(3): 752-758.]

基于大数据方法的垃圾焚烧发电厂环境执法监管数据预测模型探究

薛军¹, 奚强², 徐淑民¹, 张宏伟^{1,✉}, 桑宇¹, 乔鹏¹, 侯鑫³, 刘占上³

1. 生态环境部固体废物与化学品管理技术中心, 北京 100029; 2. 绿色动力环保集团股份有限公司, 深圳 518057; 3. 北京市保生源科技有限公司, 北京 100080

摘要 在垃圾焚烧发电厂运行系统负荷特性统计指标和污染源监督性监测数据积累的大数据背景下, 有效提取数据之间的关联特征对于垃圾焚烧系统规划运行和执法监管具有重大意义。首先, 通过 Pearson 关联分析获得运行负荷特性指标和排放特征指标任意 2 因素之间的相关性特征; 然后利用 SPSS 及 Python 软件, 构建用以预测烟气污染因子排放量的多元线性回归模型及 BP 神经网络模型。对模型预测结果进行了比较, 结果表明, 多元线性回归模型和 BP 神经网络模型都能应用于烟气污染因子排放量的预测, 进一步得出的 BP 神经网络模型的预测效果优于多元线性回归模型。本研究对于探究工业污染源环境执法建模和定量分析污染源排放水平具有参考价值。

关键词 垃圾焚烧; 大数据; 运行负荷特性指标; 相关性; 多元回归分析; BP 神经网络

生态环境保护工作在经历了生态保护、法律手段和市场机制之后, 已进入以创新为驱动力的全新环境治理时代。十三届全国人大常委会第十七次会议审议通过了修订后的《中华人民共和国固体废物污染环境防治法》^[1], 该法自 2020 年 9 月 1 日起施行, 其第五十六条规定: 生活垃圾处理单位应当按照国家有关规定, 安装使用监测设备, 实时监测污染物的排放情况, 将污染排放数据实时公开。监测设备应当与所在地生态环境主管部门的监控设备联网。在新冠肺炎疫情进入常态化防控阶段的情况下, 中国生态环境执法坚持“方向不变, 力度不减”, 通过“非现场”监管方式, 转变执法方式、执法途径并提升执法效能, 成为重要的监管创新手段。

随着中国环境保护进程的不断深化, 环境监管过程中产生的监控大数据日趋丰富, 而有效获取企业生产数据和监管数据中潜在的信息和知识, 对于指导垃圾焚烧系统规范运行和执法监管具有重大意义。多年来, 垃圾焚烧发电企业已经积累了巨大的生产运行负荷特性统计数据 and 污染源排放数据(排放特征指标), 数据的数量和质量都在不断提升。经验表明, 运行负荷特性指标已经成为影响焚烧系统排放特征指标的重要因素。而且, 生产技术装备水平和环保技术装备水平的提高对排放指标的影响日益加重。环境主管部门在环境监督管理过程中, 通过探究生产过程中的运行负荷指标与污染物排放特征因子间存在的关系, 可以通过部分指标的变化推测其他指标的变化

收稿日期: 2021-12-30; 录用日期: 2022-03-15

基金项目: 国家重点研发计划“固废资源化”重点专项 2019 年度定向项目(2019YFC1904600)

第一作者: 薛军(1973—), 男, 博士, 副研究员, xuejun@meescc.cn; ✉通信作者: 张宏伟(1985—), 男, 硕士, 工程师, zhanghongwei@meescc.cn

特征及负荷特性数据的异常^[2]。

目前，模型预测技术主要包括数值模式方法和统计与机器学习方法^[3]。数值模式在大气污染物预测上应用较广并取得了一定的成就，主要使用各类气象数据和排放源数据等，通过大气污染物扩散和物质的物理化学过程来模拟污染物的形成^[4-5]。近年来，随着机器学习的发展，许多研究开始采用机器学习技术进行模型研究和预测^[6]。这些研究主要集中在宏观空气质量预测、处理技术^[7]，而对垃圾焚烧发电厂污染源排放的预测评估研究相对缺乏。

本研究结合垃圾焚烧发电行业专项执法行动中的应用实践，使用垃圾焚烧发电厂的物联网监管数据和污染源自动监控数据，构建多元回归统计模型和BP神经网络(BPNN)机器学习预测模型，对某垃圾焚烧发电企业的污染排放数据进行预测，并分析不同模型的预测精度。拟通过对比不同模型的预测结果，建立适用于我国垃圾焚烧发电厂环境监管执法的预测模型。

1 材料与方法

1.1 研究对象和数据源

目前，垃圾焚烧的技术流派主要包括炉排炉工艺和循环流化床工艺2种，国内有40%的垃圾焚烧企业采用循环流化床工艺。循环流化床工艺是我国自主开发的炉型，在我国拥有完全自主知识产权。在实际运行过程中和历次监督检查中，循环流化床垃圾焚烧炉超标比例高于炉排炉。选取国内3条循环流化床焚烧炉生产线(分别记为A、B、C)用于研究，在焚烧厂的生产装置上加装了监控传感器，经过物联网长期检测各个点的采样数据，形成了一定时间阶段内，针对各个监测点位的时间序列数据。

1.2 数据预处理

由于各类监测数据存在异常值或缺测值等情况，需要对各类原始数据进行数据清洗，主要包括：1)对于传感器监测的历史数据等原始数据，均需剔除其中的异常值和缺失值，且如果传感器数据中的某一项缺失或者异常，则该小时所有数据全部剔除；2)对于PM_{2.5}浓度数据，在剔除原始数据异常值的同时，对于一行记录的所有数据项都是“0”或整个列的数据没有变化的数据也进行剔除处理。选择3个独立运行生产线，在2019年4个月时间内内采样点的传感器每5 min采样数据，通过数据质量控制后，最终选用数据共26 780条。在数据整理后，进行数据集划分，其中训练集21 424条(80%)，测试集5 356条(20%)。

1.3 模型构建

首先进行相关性评价，根据评价指标来获取各模型的最佳参数，并使用最佳参数和训练集进行各机器学习模型的构建。最后将训练好的模型用于预测。

1)运行负荷指标与污染物排放特征因子相关性建模。选取运行负荷指标与污染物排放特征因子作为研究变量，其中，解释变量分别记为：给煤机、1#给料机、2#给料机、二次风机、灰尘输送机、1#活性炭风机、2#活性炭风机、1#喷石灰粉罗茨风机、引风机、1#活性炭罗茨风机、2#活性炭罗茨风机、干粉风机、流化风机的输出功率、布袋除尘压差、炉顶点位1温度、炉顶点位2温度、炉顶点位3温度、炉断层点位1温度、炉断层点位2温度、炉断层点位3温度、烟气氧含量。被解释变量分别为：出口二氧化硫质量浓度、出口氮氧化物质量浓度、出口一氧化碳质量浓度、出口氯化氢质量浓度。

设 v_i 和 v_k 为其中任意2个变量，则其皮尔森(Pearson)相关系数 ρ_{ik} 计算方法见式(1)。

$$\rho_{ik} = \frac{\sum_{j=1}^m (v_{ij} - \bar{v}_i) \cdot (v_{kj} - \bar{v}_k)}{\sqrt{\sum_{j=1}^m (v_{ij} - \bar{v}_i)^2 \cdot \sum_{j=1}^m (v_{kj} - \bar{v}_k)^2}} \quad (1)$$

式中： v_{ij} 为第*i*个变量在*j*月份序列的值； v_i 为第*i*个变量序列的均值； v_{kj} 为第*k*个变量第*j*月份序列的值； v_k 为第*k*个变量序列的均值。若 $\rho_{ik}>0$ 则变量之间正相关， $\rho_{ik}=1$ 则完全正相关， $\rho_{ik}<0$ 则变量之间负相关， $\rho_{ik}=-1$ 则完全负相关， $\rho_{ik}=0$ 则完全独立，可以去除。皮尔森相关系数定量描述了2因素之间的线性相关紧密程度，可在[0,1]区间上将因素相关系数进行强弱区间划分^[8]。

2) 运行负荷指标与污染物排放特征因子的多元回归分析。结合物理意义和相关系数计算可以获取2因素之间线性相关紧密程度，但无法获取多个变量潜在的动态关系。而决策者非常期望获得不同工况影响下各运行负荷指标与污染物排放特征因子多变量所存在的各种动态关系，或期望获得特定特定时间指标之间复杂的物理勾稽关系的定量模型。数值模式方法如多元回归模型、灰色关联模型等就是通过利用多个指标的历史值进行回归分析来判别其中某一指标的变动规律，研究表明，其更加接近实际，且已在多个研究领域得到广泛应用^[9]。根据研究需求，本研究使用多元线性回归模型开展估测实验。以污染物排放特征因子序列数据为因变量(目标变量)，以运行负荷行为特征序列数据等为自变量(解释变量)，通过多元回归分析构建模型，从而对排放量进行估算。设样本总数为*n*，排放量实测值用向量表示为 $Y_{n \times 1}$ ，对应解释变量维数为*m*，用 $X_{n \times m}$ 来表示自变量因子， $\beta_{m \times 1}$ 为未知参数向量， ε 为误差项，其均值为0且方差大于0，并假定 $\varepsilon \sim N(0, \sigma^2)$ 。多元线性回归模型的矩阵表示为： $Y = X\beta + \varepsilon$ 。借助Python语言构建模型。

3) BP神经网络模型。BP神经网络(BPNN)是一种按照误差逆向传播算法训练的多层前馈神经网络,在解决非线性系统问题时,优势非常明显^[10],常被广泛应用于预测、分类、模式识别和聚类等领域。BP神经网络模型通常由输入层、隐含层和输出层组成。在模型构建中,隐藏层层数、神经元个数均根据经验公式和神经网络设计原则进行设计^[11]。3层BP神经网络如果有足够多的隐藏节点,可以逼近任何函数。

1.4 分析方法

本研究采用决定系数 R^2 、均方差RMSE、平均绝对误差MAE这3个指标对模型进行评估,各指标计算方法见式(2)~式(4)。

$$R^2 = \frac{\left(\sum_{t=1}^n (y_{o_t} - \bar{y}_o) \cdot (y_{m_t} - \bar{y}_m) \right)^2}{\sum_{t=1}^n (y_{o_t} - \bar{y}_o)^2 \cdot \sum_{t=1}^n (y_{m_t} - \bar{y}_m)^2} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_{o_t} - y_{m_t})^2} \quad (3)$$

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_{o_t} - y_{m_t}| \quad (4)$$

式中：*n*为为数据的数量； y_m 为预测结果； y_o 为真实值； \bar{y}_o 和 \bar{y}_m 分别为预测结果和真实结果的平均值。

2 结果与讨论

2.1 污染物排放时间浓度特征

以垃圾焚烧生产线C为例,将4个月采样周期内采样点的传感器每5分钟采样数据的浓度数值进行算数平均,得到不同排放指标日均浓度变化特征(图1)。污染因子排放浓度随时间的变化规律存在较大的差异性,这也说明了垃圾焚烧过程的复杂性。运用合适的炉膛和炉排结构,使垃圾在焚烧炉中得以充分燃烧;控制焚烧炉内烟气出口温度不低于850℃,烟气在炉膛及二次燃烧室内的停留时间不小于2s, O_2 的质量浓度不少于6%,并合理控制助燃空气的风量、温度和注入位

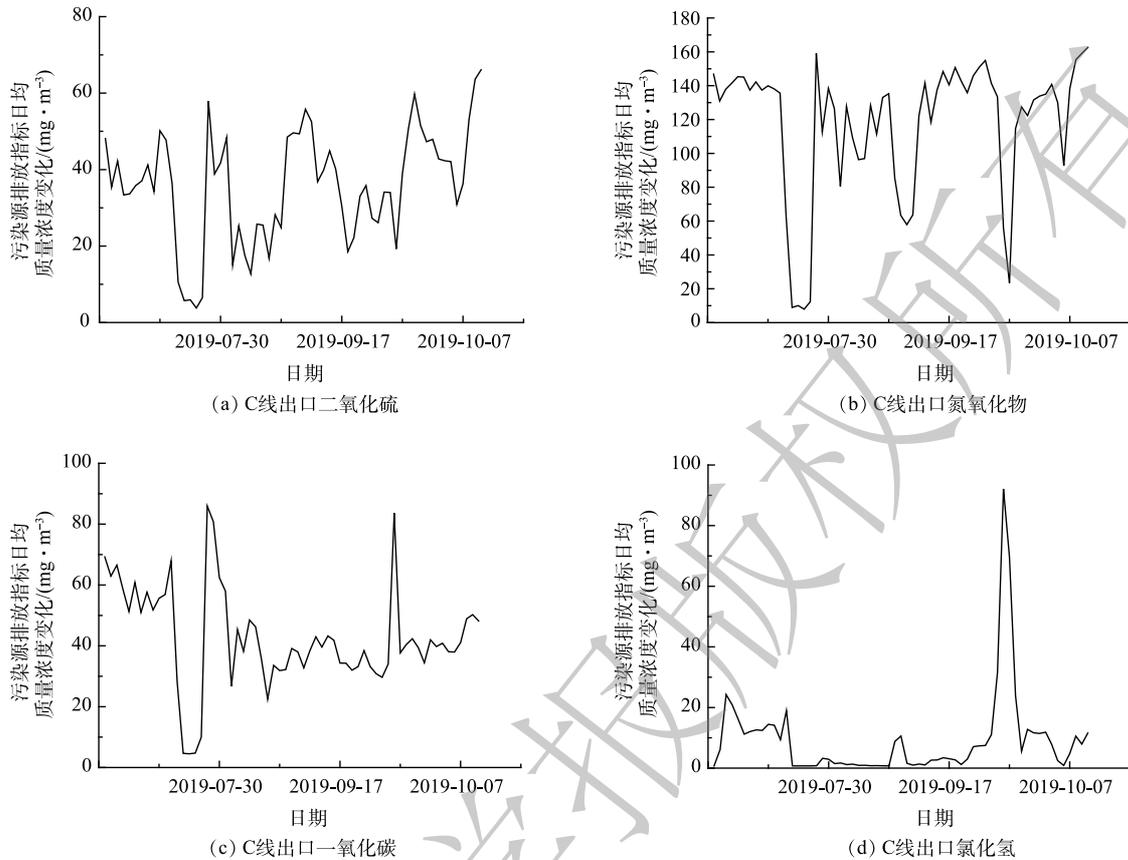


图 1 垃圾焚烧生产线 C 污染源排放指标日均质量浓度变化

Fig. 1 Daily average concentration change of emission index of garbage incineration line C

置，使焚烧系统在额定工况下运行，这些措施都有助于原始排放物浓度降到最低。衡量垃圾是否充分燃烧的指标之一是烟气中 CO 的质量浓度，CO 质量浓度越低说明燃烧越充分。排放指标变化规律不稳定，则说明垃圾焚烧过程处于非稳态系统。

2.2 行为特征序列和一系列关联因子序列的多元回归分析建模分析

首先进行垃圾焚烧排放数据的多元相关分析，将上述数据进行 Pearson 关联度的矩阵计算，可以得到关联矩阵。关联矩阵法是常用的系统综合评价法，它主要是用矩阵形式来表示每个替代方案有关评价指标及其重要度和方案关于具体指标的价值评定量之间的关系，通过将多目标问题分解为两指标的重要度对比，使评价过程简化、清晰。从相关系数结果来分析，排放因子 CO、NO_x、SO₂、HCl 与炉顶温度、炉断层温度、给料机、二次风机、给煤机流量以及布袋除尘压差具有较强的相关性。

通过相关性分析得出，排放因子 CO、NO_x、SO₂、HCl 与运行负荷指标变量之间存在线性关系，为了更准确评估各变量与排放因子之间的数量关系，以 SO₂ 为例进行计算，借助 Python 语言进行目标变量排放因子与运行负荷指标等多解释变量的多元回归分析，评估各变量对排放因子的影响。通过进行同类均值化处理和降维处理，得到 3 条垃圾焚烧生产线的线性回归模型表达式见式 (5)~式 (7)。

$$Y_A = -0.014 \times X_1 - 0.005 \times X_2 - 0.017 \times X_3 + 26.573 \tag{5}$$

$$Y_B = -0.009 \times X_1 - 0.011 \times X_2 + 0.107 \times X_3 + 17.794 \tag{6}$$

$$Y_C = 0.04 \times X_1 + 0.77 \times X_2 + 65.86 \tag{7}$$

式中： Y_i 表示排放因子CO、NO_x、SO₂、HCl； X_1 、 X_2 、 X_3 、 X_4 分别表示降维后的自变量。

将测试集输入模型中进行计算，预测数据与实际数据的对比情况如图2所示。采用拟合优度检验法对模型进行检验，Multiple R-Square为决定系数，Adjusted R-Square为可调整的拟合优度，其值越大表示解释变量对目标变量的解释程度和方程的拟合优度越高。C生产线模型决定系数 R^2 为0.49， F 统计量值为23.164($P < 0.01$)，这表明模型的拟合效果一般。

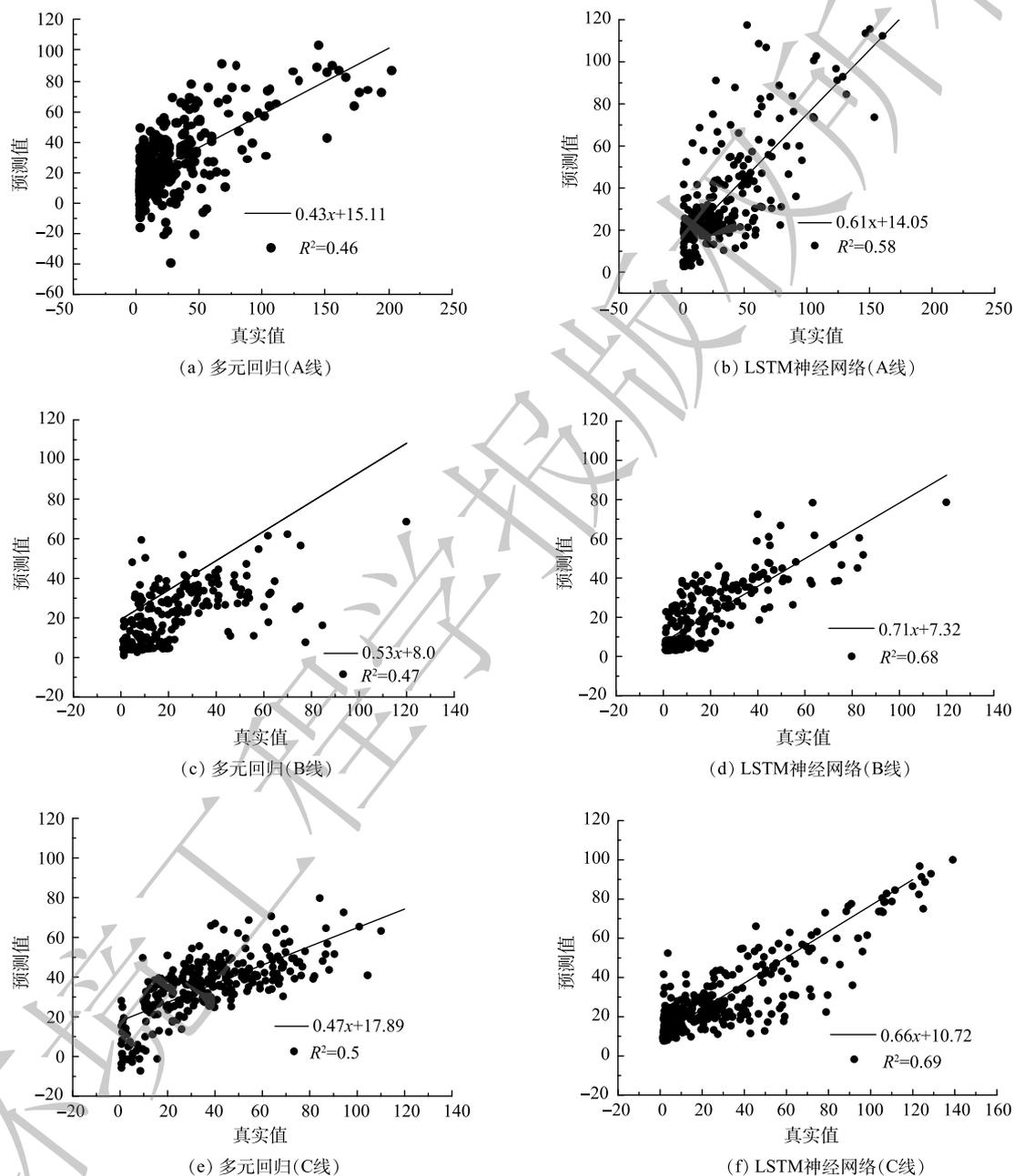


图2 不同模型的预测值与实测值的对比

Fig. 2 Comparison of prediction results and observation values under different models

2.3 LSTM神经网络模型建模分析

在Python平台下采用3层LSTM神经网络，输入层的神经元个数为与出口SO₂相关的入口参数，共22个；隐层神经元个数为60，输出层神经元个数为1，对应排放因子SO₂实测值，误差采

用均方误差。通过 800 次训练，选取损失值最小条件下的模型，并求出模型在测试集上的 R^2 、RMSE、MAE。

用 A、B、C 生产线数据建立的多元线性回归模型与 BP 神经网络模型，模拟值与实测值之间的拟合度 (R^2) 相比较而言，BP 神经网络模型具有较高的拟合优度及良好的建模效果。

通过计算实测值和预测值的 R^2 、RMSE、MAE 等指标 (如表 1 和图 2 所示)，A、B、C 线的多元回归模型的 R^2 处于 0.37~0.49，结果表现不佳。A、B、C 生产线的 LSTM 神经网络模型的 R^2 在 0.46 与 0.84 之间，结果较为理想。B 生产线多元回归模型的 RMSE 值最低，而 LSTM 神经网络模型的 RMSE 值最大达到了 34.41。对比 MAE 值，B 生产线多元回归模型的 MAE 值最低为 9.15，MAE 最大值为 C 线 LSTM 模型，值为 22.25。综合对比 6 个模型，C 生产线的 LSTM 模型的预测性能最好，而 A 生产线的多元回归模型在 6 个模型中的预测能力较差。

由表 1 可看出，通过对比分析，构建的污染因子排放量多元线性回归模型及 BP 神经网络估算模型的平均预测精度分别为 50.07% 和 71.65%，BP 神经网络模型在 3 个精度评价指标上都高于多元线性回归模型。这表明，运用 BP 神经网络模型估测垃圾焚烧发电厂烟气污染物排放指标优于多元线性回归模型，能达到更好的拟合效果。这与冯鑫伟等^[12]、徐辉潘等^[13]研究结论相一致。

表 1 不同垃圾焚烧生产线模型预测结果

Table 1 Forecast results of different garbage incineration production lines

模型	A			B			C		
	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE
多元回归	0.37	17.17	27.29	0.45	9.15	13.19	0.49	12.92	19.10
LSTM	0.54	18.71	25.31	0.46	11.29	15.62	0.84	22.25	34.41

总体而言，LSTM 神经网络模型表现最稳定，该模型的 C 生产线预测结果为所有预测结果中最好的。预测结果的差异除了与不同机器算法的特性相关外，还与不同焚烧炉生产线的工况因素息息相关。

由于因变量 (污染物排放量) 与其影响因素之间并非完全典型的线性相关关系，同时，污染物排放量的各个影响因素之间具有一定的相关关系，多元线性回归模型无法准确地描述这种非线性相关关系，因而会造成一定的偏差。而 BP 神经网络模型属于隐含式的数学方法，在分析过程中不需要建立数学模型，只需要通过原始数据训练得到稳定的学习网络，进而对数据进行预测，这种方法更加简便快捷。因此，神经网络模型可以更好地模拟非线性方程，污染物排放量与其影响因素的非完全线性相关关系以及各个影响因素内部之间的相关关系不会对神经网络的预测产生影响，进而在一定程度上提高了神经网络的预测精度^[12]。

目前，在环境管理领域，学者们利用模型进行大气污染等预测评估时多数采用多元线性回归模型，其精度差异亦较大。这是因为，不同研究领域、不同数据类型的差异会对精度产生一定的影响，而且选择不同的自变量因子对模型精度的影响也不相同。有些同样主题的研究得到的多元线性回归模型或神经网络模型的预测精度差异也较大^[13]，造成这种差异的原因与选择的自变量因子有关。另外，本研究还存在继续深入的空间，如本研究使用了 5 min 采样数据浓度数值进行算数平均，得到不同排放指标日均浓度变化特征来参与计算，按照经验也可以考虑使用中位数来进行表征，其效果可以在后续工作中做进一步的对比研究。

3 结论

1) 选取炉顶温度、炉断层温度、给料机、二次风机、给煤机流量以及布袋除尘压差等指标作为影响污染物排放特征因子的指标因素，运用物联网监管数据和污染源自动监控数据建立了预测

污染物排放特征因子的BP神经网络模型和多元线性回归模型,计算结果表明,2种模型均能很好地应用于污染物排放特征因子的预测。

2)BP神经网络的非线性映射关系对变量的处理效果优于线性回归模型的处理效果。因此,神经网络在污染物排放量的预测中将得到更广泛的应用,可为环境监管在线执法预警提供决策参考。

参考文献

- [1] 新华社. 中华人民共和国固体废物污染环境防治法[EB/OL]. [2020-4-30]. http://www.gov.cn/xinwen/2020-04/30/content_5507561.htm.
- [2] 马瑞,周谢,彭舟,等.考虑气温因素的负荷特性统计指标关联特征数据挖掘[J].中国电机工程学报,2015,35(1):43-51.
- [3] 孙宝磊.基于BP神经网络的大气污染物浓度预测[D].昆明:昆明理工大学,2017.
- [4] 李锋,朱彬,安俊岭,等.2013年12月初长江三角洲及周边地区重霾污染的数值模拟[J].中国环境科学,2015,35(7):1965-1974.
- [5] 周广强,谢英,吴剑斌,等.基于WRF-Chem模式的华东区域PM_{2.5}预报及偏差原因[J].中国环境科学,2016,36(8):2251-2259.
- [6] 康俊锋,黄烈星,张春艳,等.多机器学习模型下逐小时PM_{2.5}预测及对比分析[J].中国环境科学,2020,40(5):1895-1905.
- [7] 皮冬勤,陈焕盛,魏巍,等.京津冀一次重污染过程的成因和来源[J].中国环境科学,2019,39(5):1899-1908.
- [8] 石东源,蔡德福,陈金富,等.计及输入变量相关性的半不变量法概率潮流计算[J].中国电机工程学报,2012,32(28):104-113.
- [9] 付倩妮.基于多元线性回归的雾霾预测方法研究[J].计算机科学,2016,43(S1):526-528.
- [10] 威德虎,康继昌.BP神经网络的设计[J].计算机工程与设计,1998,19(2):47-49.
- [11] Lecun Y, Boser B E, Denker J S, et al. Handwritten digit recognition with a back-propagation network[J]. Neural Information Processing Systems, 1989: 396-404.
- [12] 冯鑫伟,黄领梅,沈冰.多元线性回归与BP神经网络模型在次洪量预测中的对比研究[J].水资源与水工程学报,2017,28(3):123-133.
- [13] 徐辉潘,萍宁,金魁,等.多元线性回归与神经网络模型在森林地上生物量遥感估测中的应用[J].东北林业大学学报,2018,46(1):63-67.

(责任编辑:金曙光)

Prediction model analysis of environmental law enforcement supervision of waste incineration power plant based on big data method

XUE Jun¹, XI Qiang², XU Shumin¹, ZHANG Hongwei^{1,*}, SANG Yu¹, QIAO Peng¹, HOU Xin³, LIU Zhanshang³

1. Solid Waste and Chemicals Management Center, Ministry of Ecology and Environment, Beijing 100029, China; 2. Dynagreen Environmental Protection Group Co., Ltd, Shenzhen 518057, China; 3. Beijing Baoshengyuan Science and Technology Co., Ltd, Beijing 100080, China

*Corresponding author, E-mail: zhanghongwei@meescc.cn

Abstract In the context of the statistical characteristics of the load characteristics statistical indexes of the operating system of the waste incineration power plant and the monitoring data of the pollution source monitoring, the effective extraction of the correlation characteristics between the data is of great significance for the planning and operation of the waste incineration system and law enforcement supervision. This study presented a mining method which could extract the operating load characteristic index (influencing factors) influence on the monitoring index was proposed. First, the qualitative analyses of potential physical relations among the indexes as well as quantitative calculation via Pearson correlation coefficient of historical data were coordinated to draw the correlation features between two factors. Then combined with SPSS and Python software, multivariate linear regression model and BP neural network model for the prediction of Flue gas pollution factor emissions were built. The prediction results of the two models showed that multivariate linear regression model and BP neural network model can be applied to flood volume prediction, and the BP neural network model was superior to multivariate linear regression mode. This study can provide decision-making basis for the environmental law enforcement modeling of industrial pollution source and quantitative analysis of pollution source emission levels.

Keywords waste incineration; big data; operating load characteristic index; correlation; multiple regression analysis; BP neural network